Interaction Frames for Artificial Agents

Interaction frames for artificial agents

Michael Rovatsos*
rovatsos@informatik.tu-muenchen.de
Institut für Informatik, Technische Universität München
80290 München, Germany

Research Report FKI-244-01
February 2001

Abstract

This paper introduces *interaction frames* as a novel method of "socially informed" reasoning in multiagent systems. We show that interactionist sociological theories provide valuable concepts and methods when it comes to modelling the "social level of representation" in artificial intelligence by operationalising "frames" as a generic data structure that can be used to capture interaction knowledge and by presenting a "framing" agent architecture that is based on interaction frames. We claim that, although further steps are necessary in order to develop actual implementations out of this abstract architecture, it provides an interesting perspective for a new kind of "socially autonomous" agent-based systems.

1 Introduction

In this paper, we propose *interaction frames* as a novel method for capturing interaction knowledge in artificial societies. Located at the *meso-level* of societal processes, they provide a means of representing the modalities of interdependences among agents by abstracting from (1) individuals and (2) trajectories of actions. This is achieved by using *roles* as abstractions of agents and *trajectories* as abstractions of behaviours. Further, *relationships* can be used to encapsulate more general levels of agent interdependence and *contextual knowledge* is included to manage the activation and de-activation of particular interaction frames. The process of activating and de-activating frames in particular situations called *framing* not only provides a powerful means for coordination, it can also be used to describe *structure* and *conflict* in agent societies. Finally, meta-level knowledge about relationships between frames, about the history of frames and about frame-related beliefs can be used to allow for (1) scaling from face-to-face interactions to groups, organizations and whole societies, (2) flexibly managing emergent changes in frames and (3) managing the "frame interfaces" among heterogenous agents.

Our theory is mainly inspired by sociological theories stemming from the tradition of *symbolic interactionism* and especially from the work on frames and framing done by Erving Goffman [10]. Among interactionist theories, which very nicely relate social action

*The author is supported by DFG grant Br609/11-1 under priority program "Socionics" (SPP 1077)

to the cognitive processes of individuals, Goffman's terms exhibit the specialty that they enable the evolution of structure *within* interactional processes. This has two advantages compared to (i) structuralist sociology and (ii) other interactionist theories, because it neither assumes the existence of some kind of "social structure" beyond the individual (which somehow determines the individual's actions) but still allows individuals to *make use* of structures in order to facilitate routinised interaction. Relying on such routine renders constant strategic planning unnecessary, not every move has to be an *ad hoc* improvisation/exploration (which would imply a huge waste of resources in everyday life).

The remaining sections are structured as follows: in section 2 we discuss our general outlook on multiagent societies, clarify the goals of our research and briefly discuss our stance towards the problem of "structure vs. adaptation". Section 3 introduces interaction frames as a fairly abstract data structure, discusses principles of the underlying socilogical theory and presents some simple examples. A detailed exposition of frames and framing is then given in sections 4 and 5, and the final section rounds up with some conclusions.

2 Assumptions

Before giving a detailed exposition of the concepts we have developed, we review the most important assumptions that are underlying our view of multiagent systems.

2.1 Artificial societies

Our outlook on a world inhabited by artificial agents is mainly one in which *situated*, *proactive* and *rational* (in the sense of *goal-oriented*) computational agents have to operate, whose perception and welfare depends on an environment on which they can effect changes by performing actions. We further assume that the outcomes of each agent somehow depend on the actions of other agents, and that, in general, the complexity of these dependencies (and the complexity of the behaviour of the non-agent environment) is too great for an agent to compute optimal plans in any situation within reasonable time. They therefore need to find *satisficing* ways to pursue their goals, they need to adapt to a world which is dynamic to them, and to reason about the contingencies that may arise.

Basically, we adhere to the classic *transducer* model in modeling agents and the world formally. Thus, an agent a_i is a function

$$a_i : P_i \times S_i \to S_i \times A_i$$

that will transform, at any point, a percept $p \in P_i$ into an action $\alpha \in A_i$ given some internal state $s \in S_i$ which may also change during this process. Provided that P_i, S_i and A_i are finite, such an agent is nothing but a finite-state machine.

Likewise, an environment E is nothing but another automaton, processing agent actions and generating agent percepts for a population $\mathcal{A} = \{a_1, \ldots, a_n\}$ while changing its internal state $s_E \in S_E$ (for simplicity, we omit the details of introducing further functions to describe what happens if the population changes):

$$E : (\times_{i=1}^{n} A_i) \times S_E \to S_E \times (\times_{i=1}^{n} P_i)$$

To obtain more compact representations of the state space (without extending the expressiveness of the model), we introduce a first-order-logic-like language L_i with predicate, function and variable symbols and the usual operators/quantifiers which is used by agent a_i to encode its internal belief states.

If we also introduce some discrete time-scale $T = \{t_0, t_1, t_2, \ldots\}$, we obtain the following definition for an agent

$$a_i : P_i \times 2^{L_i} \times \{t_i\} \rightarrow 2^{L_i} \times A_i \times \{t_{i+1}\}, \quad i \geq 0$$

Likewise, we can modify the definition of the environment to obtain

$$E : (\times_{i=1}^{n} A_i) \times 2^{L_E} \times \{t_i\} \rightarrow 2^{L_E} \times (\times_{i=1}^{n} P_i) \times \{t_{i+1}\}, \quad i \geq 0$$

if L_E is some language in which we can describe the facts $F_E \in L_E$ that hold in the world at t_i/t_{i+1}. For the time being, we need not make further assumptions about the type of belief/fact language we are using.

As concerns goal-orientation, we define some "quality" on beliefs by a partial order of preference $\prec \subseteq 2^{L_i} \times 2^{L_i}$ which is given *a priori*, i.e. it reflects some "bodily" ordering on the desirability of certain beliefs (desirability of percepts can be introduced by trivially extending the language of beliefs). Now since the function a_i is given, there is not much we could call "deliberation" to assert that an agent is actively pursuing its goal – only if we review its activity *a posteriori* and observe an increase in "desirability of states encountered" can we infer that it made the "right" choices.

Next, we need to make some remarks about how we view interactions. Mathematically speaking, an interaction is a *trajectory* of actions of several agents. We can define an index set $I \subseteq \{1, \ldots, n\}$ for the agents under observation (such that a_i is in that set iff $i \in I$) and define such a trajectory as a finite sequence

$$(A_{I_1}, t_{j_1}), (A_{I_2}, t_{j_2}), \ldots (A_{I_m}, t_{j_m})$$

with $A_{I_k} \in \times_{i \in I} A_i$ as joint actions of the indexed subset of agents, and $j_1 \leq \ldots \leq j_m$ indices of some ordered (not necessarily direct-successor) time-stamps. To make this more general so that we can formalise interactions among groups of agents (organisations, societies etc.), we can group agents together and also observe sequences of such group interactions by simply replacing the A_{I_k} themselves by *sets* of such joint actions $\mathcal{A}_{\mathcal{I}_k}$ with $\mathcal{I} \subseteq 2^{\{1,\ldots,n\}}$ and the sets of joint actions defined respectively.

However, these "system-level observations" of interactions do not quite capture the interdependence between agents, i.e. the effects that actions have on others' perception, beliefs and subsequent actions. So, to the least, we would have to add to the joint actions used above the current joint belief states of agents and the joint percepts they are currently processing (defined in a similar way as joint actions). We then obtain something like

$$(A_{I_1}, B_{I_1}, P_{I_1}, t_{j_1}), (A_{I_2}, B_{I_2}, P_{I_2}, t_{j_2}), \ldots (A_{I_m}, B_{I_m}, P_{1_m}, t_{j_m})$$

for the case of inter-agent interactions, and something similar for inter-group interactions[1].

At the most general level, these are the kinds of processes that we are interested in, and all that follows is an attempt to formalise, at more specific levels, knowledge about such interactions which may assist agents in managing them effectively. This knowledge will be stored in mental representations of so-called *frames* which will be included in the belief sets B_i and may help to guide agents' decision-making. It should be remarked that although we will *not* define each of the concepts to follow in terms of this formal model, we attempt to make such a formalisation possible in principle[2].

[1] An even more generic view would allow the I_k or $\mathcal{I}_k$ sets to change between steps in the trajectories, and also to consider different I_k in actions, beliefs and percepts. Due to the confusing indexing, we omit the technicalities.

[2] It should be kept in mind that what we are doing here is, in an epistemic sense, to endow the systems

2.2 A note on structure vs. adaptation

In principle, there are two possible ways of looking at knowledge in AI contexts: it might be *given* knowledge imparted to agents at design time, or *acquired* knowledge which is constructed by agents themselves during execution (of course, it might also be a mixture of both). Consequently, this also applies to the interaction frames we propose: as knowledge structures, they can be used in both ways.

From a design perspective, we might use frames in a prescriptive way, so that they are fully specified control structures which agents need to comply with. In the strictest sense, this would mean forcing interactants to perform certain joint actions under certain conditions. This strongly contradicts the sociological interactionist position, which rejects social determinism and emphasises that *all* knowledge about interactions is *solely* acquired through experienced interactions. Such a use of the concept would also suggest that we *understand* the complexity of interaction and are able to plan it at design time optimally, a view which approaches "centralised control" and is counter-intuitive from a DAI perspective. Also, it might possibly lead to a definition of frames as *protocols* which formalise procedures of interaction but fail to capture the more qualitative aspects of interactions.

Hence, an "adaptation" approach seems more adequate, but how can we propose algorithms for learning without knowing the target function of such learning? The methodological stance we assume will be to proceed in four steps:

1. Observe some aspect of agent operation that is socially relevant.
2. Operationalise this aspect and incorporate it in the interaction frame data structures.
3. Validate the usefulness of the concept through simulations, in which it is used in a prescriptive way by agents.
4. If it proves useful, develop methods which allow the concept to be acquired dynamically.

Thus, we can use our preliminary understanding of some concept (often, as for all concepts described in this paper, derived from sociological theories) to embed knowledge about it in our formalisms and conduct tests to verify whether it does improve the effectiveness of the system in which instances of the concept are used in a "hard-wired" way. To get rid of the prescriptive nature of implementing "laws", we can finally suggest (and evaluate) learning algorithms to allow the use of the concept in an adaptive, domain-independent and flexible way.

The reason for emphasising this issue and for developing such a method of proceeding is that we have found that, for many of the sociological concepts we investigated, there is an inherent tension between descriptive and prescriptive stances towards the concept (as will be discussed in the sections on roles and trajectories in more detail). Sociologists seem to be constantly "trapped" in the dilemma of talking about a society which already exists, and from which the individual is unable to "escape" and the need to explain how such a society comes about, knowing that society is man-made. Without intending to further discuss this problem here (aspects of which are often appearing under the names of "micro-macro problem", "autonomy dilemma", "hen-egg problem", "emergence debate" and others), we stress that the reader should keep this problem in the back of her head in all that follows.

---

with *our* knowledge of their behaviour. This implies that *emergent* phenomena that arise when the agents use such knowledge will perhaps not be representable in our formalism. The secret hope of our research is that, incrementally, as we refine these models, we move towards some "fixed point" of social "sense" which makes the social world completely understandable to the agents.

2.3 Goals

In the above remarks we have already used terms such "useful" and "effective" without making explicit what the goal of our efforts is. As we have stated, the formalism we propose is supposed to be adequate for capturing interaction knowledge, and in fact, our goal is to endow agents with methods to develop such knowledge, so that they can optimally manage the social issues they are confronted with. In that, we do *not*

- preclude whether such knowledge is used by designers to enforce *cooperative* behaviour or not. Agents may still be self-interested, mean or antisocial, but even such agents can use social knowledge to optimise their behaviour.
- make statements about human society. Although sociologists may (or may not) find systems built using our concepts adequate to validate the underlying social theory, this is not our primary concern. This work is *not* about social simulation.
- suggest useful social laws, coordination mechanisms or interaction protocols. We believe that these are issues subject to further investigation, while our effort is primarily thought of as a model that can be used to *facilitate* the development of these.
- concentrate on specific areas of DAI research, such as virtual communities, hybrid societies, user modeling/interface agents, believable agents, distibuted problem-solving, computational economics and market-based systems, agent-oriented software engineering or computer-supported cooperative work. Since interaction is underlying all these, we do not define a *system-level goal* by which the class of relevant systems can be identified. Instead, we suggest that our theory is useful to enhance all these *interaction systems*.

3 An introduction to frames and framing

In this section, we provide a somewhat informal overview of interaction frames as a data structure composed of various elements. We start by giving some details about the underlying sociological theory, present an overview of a version of this theory as it could be operationalised for artificial agents, and show some examples of frame instances to illustrate interesting features.

3.1 Sociological theory

Through his micro-sociological analyses of face-to-face interactions (like those discussed at length in [10]), Goffman developed the notion of *frame* to denote knowledge concerning the social world that is derived from past interactions and feeds into future enactments of those interactions. This knowledge, which may include anything from rules of communication, knowledge about social dependencies or institutional knowledge to specific domain knowledge can be used in three ways:

1. *Complexity reduction:* Being the substrate of past, similar interactions between different people, frames generalise from specific situations and can be evoked to reduce decision-making complexity in a specific situation. This makes habitual routine action possible, and the stability of such established routine is necessary for individuals to concentrate on shaping their social worlds.
2. *Structuring:* The properties of particular frames and differences/similarities/hierarchies between frames can be used to structure the social world in one's mind. As a "frame

of reference", frames provide explanations/interpretations for observed interactions by classifying, grouping, and interconnecting observed actions in social contexts.

3. *Strategic reasoning:* Knowledge about interactions can be used to achieve one's own goals. What is stressed by Goffman as the importance of "playing drama" when adhering to frames essentially means that individuals use the conventions of everyday life to further their own needs.

What are frames? As a preliminary, not very precise, definition we might propose the following:

> *A* frame *is a structure containing information about the actions, perceptions and beliefs of interactants during a certain class of interaction trajectories.* Framing *is the process of one or more parties evoking a frame in interaction encounters.*

From this definition, but also from the work done on frames in sociology, we can derive some of their characteristics to further specify the concept:

- **Interaction focus:** Frames are *always* about interaction, be it collaboration, coordination, competition, conflict, imitation, information exchange, or any other conceivable form. This means that they always "frame" the activities of several parties, and thus shape *interactions* in the first place, and not agents, groups or social structures[3] '

- **Conventional nature:** Once a frame is activated in the participating parties, it implies their commitment to comply with certain rules. These may permit some actions and forbid others, possibly also in some temporal order and in a different way for each interactant, but they may also require adherence to some "semantic space" in communication (e.g. a specific ontology) or some particular mode of negotiation. This does not, however, mean any whole-hearted acceptance of the convention, they are simply *used* by interacting individuals or groups to manage their interactions effectively.

- **Persistence:** Unlike mental representations of, for example, agent plans, a frame has its own representation which may exist quite independently from what agents actually do at a given time. A frame has its own history (which includes when it has been enacted by whom and with what success), is linked to other frames that an agent has in its knowledge base, and influences decision-making when activated, of course. This is not to say that it is some kind of blackboard or shared coordination space, but it must be explicitly represented in an agent's "mind" as "distilled experience" regardless of whether it is currently applied or not.

- **Context dependence:** The activation of a frame is always subject to certain conditions. Such a condition might be that an agent wants to carry out a certain task, an occuring unforeseen problem, but also, for example, the sudden confrontation with a resource conflict among several parties. Thus, an agent might have an "arsenal" of frames at its disposal that it can use to handle different situations, and the choice of which one to activate at the right time is a crucial issue.

- **Routine/Change and frame adherence:** By limiting the possible course of some interaction to a tractable number of "trajectories" of action, the decision space for

[3]Of course, since "somehow everything relates to everything", there is always a frame problem (in the traditional AI sense) with respect to defining "what is an interaction". We will return to this later, but, in general, we will treat any joint execution of the actions of several actors as an "interaction".

an agent is dramatically reduced, as long as the frame "works". As long as the frame serves the agent's goals, it is a "script" rigidly applied when making decisions. Conflicts arise when it doesn't suit the private purposes, when the agent is no more able to comply with it, when there is insecurity about which frame to activate, etc.

- **Shapability:** Not only may an agent propose new frames or the adjustment of existing ones, but frames may also be inherently flexible and subject to constant change. Consider, for example a customer-clerk interaction frame at the bank counter: certainly, it contains certain rules of conduct (such as signing a document when asked to do so for the customer or attempting to execute the customer's requests if possible for the clerk), but it also has many degrees of openness: How many banking transactions will occur? In which order? Who will take the initiative to continue/stop/change the interaction? What effects will external conditions have on the interactions (for example, will the interaction continue if the bank closes during the conversation – maybe, if the customer is very important for the bank, or if the conversation is critical, e.g. about a mortgage loan)?
  This means that frames may vary greatly as to the degrees of freedom they permit, and also that whatever is fixed about a frame is constantly changing through enactment and affecting the future properties of the frame.

- **Role-orientation, situation-orientation and norm-orientation:** If each individual/group had enough experience and knowledge about every other individual/group with respect to every relevant interaction, frames would just be a reformulation of that knowledge and hence not add anything to the expressiveness of agents' knowledge. Therefore, the capacity of frames to abstract from individuals and particular enactments is of paramount importance: by expressing the characteristics of certain interactions in terms of *roles* (that are instantiated with individuals or groups when the frame is activated), *situations* (that are instantiated with the current state of affairs when the frame is enacted) and *norms* (as rules that regulate the behaviours of the interacting parties) the individual is able to use more generic representations of the social world, thus better oganizing its knowledge about that world.

- **Communication-orientation:** Frames are inherently communicative. Although, in principle, it is possible that non-communicative actions, interpreted as gestures or signals of some kind, might be sufficient to frame an interaction, this is very unlikely, because non-communicative actions have usually some effect on the agent or the world (other than merely "having performed a communicative act"). This would imply that, were the frame enacted communication-less, the interaction would *be* the frame, which leaves only very restricted options to change it, to negotiate etc.
  On the other hand, this implies a *semiotic* orientation of interacting parties, because enactment of, activation of, compliance with, deviance from and changes to frames can always only be identified through a *symbol-mediated* process in which actors exchange signs without ever being sure of identical interpretation, since the understanding of the signs themselves evolves through interaction that is supposed to be regulated by using them.

These characteristics provide us with first hints to the meaning of frames and framing, which we will use in the next section to develop a possible top-level design for frames.

Attribute	*Description*
roles	Classes of agents or groups associated with expectations about skills, tasks, goals, behaviour, rights, duties, social position and status.
relationships	Interdependence between roles, i.e. relationships of dependence, control, acquaintance, authority, similarity, group membership, power.
trajectories	Statements about possible courses of interaction. These can range from simple statements about what must become true at some point to whole joint plans; they can be mandatory or negotiable (as default rules that may be overriden), positively or negatively formulated (permissions, restrictions) and possibly sanctioned.
context	Defines the situations in which the frame is relevant or not, pre- and postconditions for its activation, and also conditions that are necessary for the interaction to be sustained.
beliefs	Domain knowledge that is necessary to carry out the interaction, for example a shared ontology or communication language or meta-level knowledge that is required.
links	A list of relationships of the current frame to others (structural: subsumption, inheritance, aggregation, sequencial/parallel coupling; semantic: alternative, shared attributes, etc.).
history	A sequence of past enactments common to the interacting parties stored as a series of frame transformations and/or concrete instances.

Table 1: Common attributes of a frame.

3.2 Data structures for frames

Frames consist of two parts: (1) *common attributes*, i.e. knowledge about the interaction that is assumed by the interacting parties to be common knowledge among them (roles, rules, relationships, contexts, beliefs, frame history and links to other frames) and (2) *private attributes* which supplement the data structure with information that is of interest to the agent/group while applying the frame (role assignments, activation status, frame tracking information, evaluation information).

Thus, while common attributes are used to manage the shared knowledge about frames as objects, private attributes can be used to manage instances of frames in ongoing interactions.

Let us now present these two blocks one after another. Table 1 shows the common attribute section.

In it, the "roles" attribute captures information about the interacting parties, as to their expected behaviour, capabilities, social standing and volitions (these will be discussed in more detail in section 4.1). This attribute must provide definitions for all those types of agents or groups of agents (including formal institutions, organisations or whole societies) participating in the interaction.

The "relationships" part encapsulates relationships between the defined roles which are relevant in the current interaction. Information about these relationships may be used to support interpretation and decision-making processes if the action rules provided in the "trajectories" section fail to provide a complete picture. For example, if a norm states that lying is sanctioned, but the lying party is more powerful than the party being lied to, the betrayed agent might refrain from applying the sanctions.

As has just been mentioned, the "trajectories" attribute defines the "rules of conduct" related to the class of interactions that the enclosing frame represents. Since the characteristics of such norms is both complex and arguable, we discuss the concept at length in section 4.2.

As opposed to "trajectories", the knowledge stored under "context" relates to the activation and de-activation conditions the frame is subject to, to rules which must hold while it is enacted and to conditions that must hold before and after the execution of the frame.

"Beliefs" summarise the assumptions about the world that must necessarily be shared by the interacting parties so that the frame can be carried out correctly. For example, if the frame is defining some sort of collaborative work, it is a prerequisite that all agents know the goals of this work and when these will be attained, or how to measure progress in fulfilling the task. In that case, the beliefs would need to capture this domain knowledge which is not directly part of the interaction but a prerequisite for performing it appropriately. Alternatively, beliefs might contain knowledge representing consensus on how to interpret certain observations and communicative acts, i.e. a shared symbolic universe.

The "links" section contains *meta-frame* information that places the current frame in the context of some hierarchical relationship to other frames, which it might enclose, of which it might be a sub-routine, a sub- or super-class, and so on. Apart from these structural relationships to other frames, there might also be semantic links, if, e.g. two frames share roles, trajectories, activation rules, beliefs, if one frame is an alternative to the other (a more specific case of frame similarity) or if it is a pre- or postcondition of some other frame (e.g. "patient-doctor consultation" frames are mostly preceeded by "patient-secretary appointment scheduling" frames).

Frame "history" data, finally, lists the experience of the agent with the frame as a series of past enactments and/or a series of transformations of previous forms of the same frame, as it has evolved over time. Although these relationships might also be included in the "links" attribute, we provide a special slot for them because (1) they do not point but to (previous versions or instances of) the same frame and (2) we regard the evolution of frames as an issue that deserves particular attention, as it reflects the dynamics of interaction over time.

As concerns private attributes, these are summarised in Table 2. Basically, they include a "status" attribute for each "common" attribute which relates the general statements about roles, trajectories, activation and beliefs given in the "common attributes" section to the specific situation the actor is in, except for the "evaluation" slot which connects the frame to the "non-interactional" reasoning mechanism and knowledge base of the agent or agents acting as a group. This is necessary to put the use of the interaction frame in the context of a rational agent or agent group (in the case of a group, this part might again have to be the "target" of some other or several other interaction frames, by which the group members negotiate the properties of such an evaluation slot).

However, the "status" attributes are not independent from each other: the correctness of role assignment must be constantly underpinned by observing the actions of others (and, too, one's own actions) but those actions must also be interpreted with respect to the existing norms as adherent or deviant actions and to revise the state of current beliefs

Attribute	*Description*
role assignment	An assignment of the roles appearing in "roles" to actual agents, groups, organisations etc. May involve tentative assignments or statements about the uncertainty of the "matching" process. Additionally, specific information about beliefs that support or contradict the current assignment may be supplied here.
trajectory status	Encapsulates information about the current status of frame execution. This includes keeping records of observed action and communication and relating this to the "trajectories" attribute, possibly enriched by information about deviance from and adherence to these norms, enforced sanctions etc. It may also include information that can be explained as a result of the relationships and not as a direct consequence of the interaction rules.
belief status	Information on the epistemic properties of the current implementation of the frame, for example, uncertainty about the degree to which allegedly common knowledge is actually shared by all interactants.
activation status	Keeps track of beliefs about the applicability of pre-conditions, post-conditions, activation and de-activation conditions and sustainment conditions. May also include statements of doubts/uncertainty concerning the relevance of this frame with respect to the ongoing interaction or qualitative statements about "degree of activation" and explanations for these.
evaluation	This slot relates the properties of current interaction to the private goals, needs and beliefs of the agent. The data contained herein is used to make decisions concerning modifications or abandonment of the frame if it doesn't serve the agent's goals. Explanations for the current assessment of the frame may be given.

Table 2: Private attributes of a frame.

of interacting parties[4].

Most importantly, all information about the interaction status stored in these attributes must be used to make decisions concerning the activation status (together with identifying the applicability of boundary activation conditions, and combined with the current evaluations of the interaction).

Taken together, these private attributes link (1) abstract interaction to concrete enactment and (2) social structure to cognitive representation, thus providing a basis for *using* frames from an agent's point of view.

Having sketched the data structure informally, we next turn to some examples to illustrate how they might be used in practice. These examples will also be kept informal, as their formal modeling requires settling on some more application-specific vocabulary.

3.3 Examples of frames

The two examples we use in this section are deliberately chosen to illustrate the varying degrees of freedom which interaction frames may leave to the interactants: first, a game of chess that is a very rigid form of interaction is presented and second, a simple negotiation protocol without fixed duration and very few restrictions is outlined.

3.3.1 Playing chess

Let us look at a game of chess between non-professionals, which is very useful to discern the different types of information included in the attributes:

Name: *Chess-frame*
Common attributes:

> **roles:** Player-White, Player-Black. Their skills are (1) moving any pawn to an arbitrary position on the board, (2) "laying down" the king to signal giving up, (3) saying "Check" and "Check-mate". It is both players' goal to check-mate the other. There is no prior expected behaviour by virtue of these roles but that Player-White will make the first move, and that the other player will wait for this to happen.
>
> **relationships:** The players are equal adversaries: they are independent from each other except with respect to what is happening on the board. Each one's gain is the other's loss.
>
> **trajectories:** The rules of a chess game as to pawn movements, termination conditions (also in "tie" situations) etc. Possibly time limitations on moves. All these rules are also norms, because they include sanctions: false moves have to be made undone, missing out on saying "Check" implies that the attacking player loses, etc.
>
> **context:** Depends on "non-chess" activities of the parties. Usually, a pre-condition is that both players declare their intention to play, that there is a chess board with pawns available, a post-condition is that the result of the game is defined after termination. Activation starts when both players agree to play, de-activation occurs upon termination of the game. Sustainment conditions include that noone quits the game by "leaving", admitting defeat, and the maximum time that each will wait for the other's move is usually bounded.

[4]This raises the interesting question of whether unexpected behaviour should be assigned to one's own faulty understanding of roles, deliberate deviance, erroneous interpretation, inconsistencies in frame activations among the parties, etc.

beliefs: Communication language including "Check" and "Check-mate" primitives understood by both players. "Correct" interpretations of board states, understanding of the rules.

links: specialises: *Two-Player-Game*, *Leisurely-Interaction-Frame*, *Non-Risk-Competitive-Game*; shares-goals-with: *Resource-Conflict-Frame*; pre-condition-frame: *Two-Party-Activity-Invitation* (etc.)

history: Frame has remained identical for all enactments.

Private attributes:

role assignment: Player-White: self, Player-Black: Tony. Full "match" concerning capabilities and behaviour so far, but Tony is my boss and I am only playing because he asked me to. I will let him win.

trajectory status: A series of moves made, current state of the board (in particular, pawns defeated by each other), anticipations concerning next n moves.

belief status: Tony believes I want to play while I don't. He also thinks I want to win, and that I will do my best, which is wrong. As concerns the game itself, the common knowledge required by the frame is ensured.

activation status: Full activation, we are both playing and none of us indicates withdrawal from the game. The game is not finished yet, everyone is adhering to the rules.

evaluation: Since I have no interest in the game, there is a conflict with my private goals (I want to go for dinner). Next time I will try to find an excuse, but I cannot leave this frame due to social pressure.

As concerns the common attributes, this is a strictly pre-defined and structured protocol-like interaction. One might argue how much of the goals, skills and behaviours of the frame should be encoded in role models, and how much of them should be declared a "norm". The important thing to note here is that it depends on whether the role "models" are going to be used in different frames, as well. For example, if we also have frames of other two-player games, even more information could be included in the role definitions (e.g. playing in turns, the right to consult the manual, etc.) These would then be other sub-classes of some imaginary super-classes *Two-Player-Game*, *Leisurely-Interaction-Frame* and *Non-Risk-Competitive-Game* the agent has constructed; the relation shares-goals-with allows us to link the frame to a frame about (serious) resource conflicts, so that, for example, if in found in a row over some resource, the agent might suggest a game of chess to resolve the issue! Another interesting point to note is the discrepancy between the role definitions of the game and the actual dependencies between the interactants in the private attibutes, induced by knowledge external to the frame. Finally, the apparent attitude of the agent in question who is pretending to like the game shows that this framework enables us to formalise the Goffmanian notion of "drama".

3.3.2 Simple negotiation

The crucial point to note about frames is that they can supply information necessary for fruitful *coordination* [7]. To illustrate the strengths of frames as a tool for capturing coordination knowledge [1], we might look at a possible model for a simple "negotiation" frame. Within this frame, the two parties are assumed to "discuss" a certain issue such that consensus is reached, and they do so by exchanging arguments. It is thereby left open, whether these arguments are logically linked to the issue in question (they might also be "external", like offering money to bribe the other), but the frame only remains active as long as each one is trying to convince the other.

Name: $Negotiation - Frame$
Common Attributes:

> **roles:** "Arguers" A and B. *Skills:* Make or retract a proposition, accept or reject any of the other's propositions. *Goals:* Reach consensus by leaving no propositions open or "denied". *Expected behaviour:* Not specified.
>
> **relationships:** None can be preassumed.
>
> **trajectories:** Talk in turns, start by tossing a coin. In each step, add or retract a proposition or deny/reject some of the other party's propositions. Logical inconsistence in statements over time is punished by canceling the frame. The same counts for inconsistence with facts in the "beliefs" attribute.
>
> **context:** Begins with any unsolicited assertion of A to B or vice versa. Frame ends iff for every proposition there exists an "accept" statement of the other party. The post-condition is that both agents are assumed to believe the statements agreed upon.
>
> **beliefs:** Both believe that neither A nor B knows whether they accept/reject each other's opinions a priori. Both believe in a closed-world assumption and in monotonicity of assertions. Common logical language and ontology.
>
> **history:** None.
>
> **links:** (Not considered here.)

Private Attributes:

> **role assignment:** ...
>
> **trajectory status:** ...
>
> **belief status:** ...
>
> **activation status:** Full activation.
>
> **evaluation:** ...

(To keep the exposition concise, we omit details about private attributes.)

This frame defines the following family of possible interactions: two parties make (in turns) a series of assertions. These can either be additions of new propositions (we call these *atoms*), retractions of prior own atoms, or accept/reject statements (which we call *evaluations*) with respect to any previous atom asserted by the counter-party. The frame terminates if all atoms on both sides have been accepted by the other (and all rejected atoms have been retracted by their owner). After such an agreement has been reached, it is assumed to be the common belief of both parties.

Notice that this frame is leaving several degrees of freedom to the agents:

- the *topic* of the argument is not defined, nor is the type of arguments that are used,
- the frame says nothing about antagonism, social sense or cooperativeness of the parties; the only alleged goal of the parties is to reach an agreement (which seems to be a reasonable assumption, otherwise they would not engage in such a dialogue),
- it doesn't make assumptions about whether they are agents, groups, organizations, or institutions.

At the same time, it is very specific in other respects:

- It does not allow agents to "change their mind": any inconsistency with previous statements, be they atoms or evaluations, will imply abandoning the frame.
- It has a very specific termination condition: if everything is agreed upon, the discussion ends. This means that once it starts with some unrequested proposition by anyone, it cannot proceed after agreement. Also, in case of disagreement, it may go on forever. (Whether this is a desired behaviour or not is a different issue.)
- It is conflict-free. Conflicting views do not imply any change of behaviour, the interactants are being completely "sobre" about the negotiation.

We believe that this nicely illustrates the richness of the frame data structure and the manifold possibilities in defining such frames, frame hierarchies and networks and the possibilities for agents to incorporate them in their reasoning processes (from blind obedience to "cheating", from conservatively sticking to established practice to constantly questioning it by deviating from expectations).

Having outlined the concepts more or less informally, we will discuss individual attributes of frames and an algorithmic view of the process of framing in the following two sections.

4 Frames

The common attributes of frames which we will discuss here in more detail (private attributes are described in section 5 - "Framing") roughly fall into three categories:

1. *information about actors and their relationships*,
2. *information about actions* and
3. *information about frames*.

In the following, we discuss possible formalizations of these three with respect to the notion of frames we have developed so far and the intuition behind it.

4.1 Abstracting from individuals: roles and relationships

The concept of a role has been widely discussed both in DAI and sociology in many contexts and with various interpretations. For our purposes, it suffices to assume that a role is some generalised model of the "other", which we define informally by the equation

$$role = expected\ behaviour + social\ position$$

thereby stressing that a role implies a certain behaviour as well as a position in a set of relationships which exist in a social context.

As concerns a role's expected behaviour, three different levels can be distinguished: a *behavioural* level at which knowledge about the can be deduced from action observation solely, an *intentional* level that includes the agent's goal goals and beliefs (which can only be communicated) and a *social* level at which the agent's actions' relationship to other agents is analysed (again, these might not be observable). Thus, we have three different levels at which the *same* activity might be described[5].

It should be kept in mind that expectations about behaviour may arise both from experience and statistical observation as well as from projected aspirations and fears

[5]In the following, we shall adhere to the "learning" and "modelling" stance when describing frame attributes. Clearly, similar implications arise in prescriptive terms, i.e. when using, for example, behaviour models for generating action rather than inferring its undelying mechanisms.

about someone's behaviour which spring from communicated intentions, own plans and anticipations, etc. This *duality* of expectations, which will hold throughout everything we say about trajectories (after all, a role is also nothing but a norm, which is actor-related rather than action-related) in the next section is a key issue, especially when it comes to learning interaction frames.

As far as relationships are concerned, these constitute the "explanatory frame" for roles since they describe the social implications of filling a role. Thus, even though, strictly speaking, they do not form part of the definition of a role, we choose to discuss them in this section, too.

4.1.1 Behavioural role attributes

Behavioural attributes are easiest to derive by observing an agent in action, since, usually, simple tracking of trajectories is sufficient to derive action expectations and information about the agent's skills.

Behaviour. The simplest form of deducing behavioural expectations is to simply store trajectories of its actions in some data structure, e.g. a probability distribution over its action options updated over time, or a learned (probabilistic) finite automaton. However, depending on the representation one chooses, difficulties may arise:

1. If we use probability tables, obviously, these neglect causality of behaviour and are reluctant to adjustment when the agent's behaviour suddenly changes. On the other hand, they are easy to create and update: unless we fail to recognise the agent's actions, our model will be consistent with observations.
2. If we assume that the agent is acting according to a stimulus-response scheme, we must (a) detect the right stimulus over time that actually caused the behaviour, (b) assume that perception is (at least to some degree) shared by observer and observee, (c) be sure about what the observed actually does. Although (c) is a basic epistemic condition that applies for any method, information about (a) and (b) is generally inaccessible and thus its modelling is a difficult undertaking.
3. If we also want to include changes in the observed agent's *state* in the expected behaviour model of the role it is fulfilling, things get worse. Then, we either have to assume a state space that is consistent with observation (as in the case of learning automata) or to rely on communication by which the agent informs us about changes in its internal state (which may be inaccurate, delayed, noisy or manipulated).

Hence, the decisions about modelling levels imply increasing assumptions, as is depicted in figure 1. Here, we can assume that, no matter which level of modelling is chosen, a general predicate for behavioural expectations of a role can be derived through observation:

$$Does(R, S, \phi)$$

to denote that any agent a_i who fulfils role R is expected to do ϕ, where ϕ is a first-order formula that contains no action symbols except those pertaining to A_i (and contains at least one such action symbol), i.e. a statement about what the modelled agent will do. S is a formula defining a situation in which the agent will do ϕ which may include preconditions that trigger ϕ. Note that although what is stated in S might simply have been included in ϕ, we require that ϕ does not make statements about others' actions while stating something about the actions of R. Also, note that if R is the role of a cluster or agents (group, organisation, population), the action symbols need to be extended to the cross-product of the individuals' action sets.

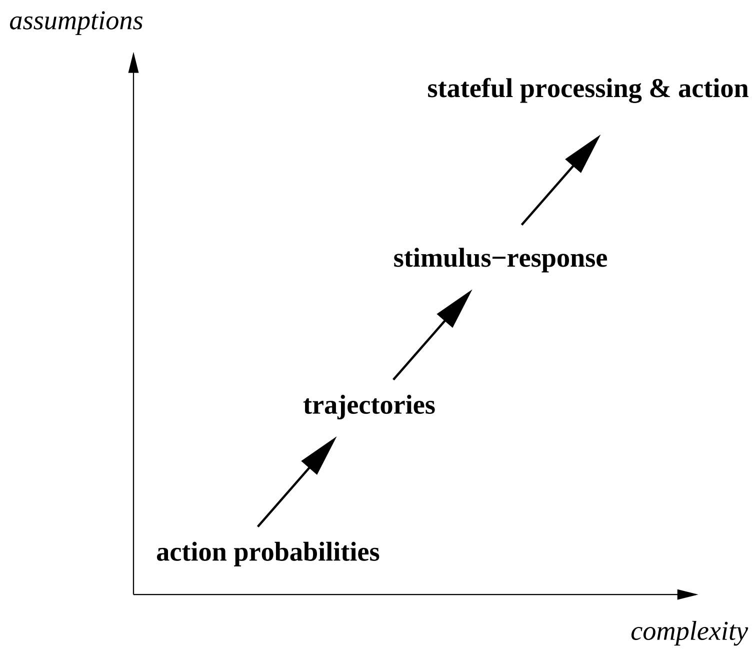

Figure 1: Modelling role expectations at different levels implies making more and more assumptions, the more expressive (complex) the expectation model is.

Although this predicate basically suffices (in terms of expressiveness) to capture role information, we will proceed to define more specific categories of agent models.

Skills. Like behaviour, skills can be derived from action observation without further knowledge about the modelled individual. In a most general sense, a skill can be seen as an agent's ability to change the state of the world (including the belief states of other agents). In order to assert that something is a skill, we have to assume a relationship of cause and effect, which might also be subject to side conditions. As a first formalisation, we can propose the predicate

$$Can(R, S_1, S_2, \phi, cond)$$

to denote that role R can transform situation S_1 to S_2 by performing ϕ (again, this formula does not contain action symbols other than those of R) if *cond* is ensured while ϕ is performed.

As before, we should consider what this formalisation precisely means: firstly, it has to be interpreted very rigidly: the model of such a skill asserts universal validity, i.e. if there is uncertainty about this skill, it will have to be broken down to "easier" sub-skills and the uncertainty has to be captured either in the pre- or postconditions or in the "sustainment conditions" expressed in *cond*. Secondly, requiring that ϕ contains but actions of R means that a skill cannot include the activity of third parties; however, these may either be part of S_1 or S_2 .

As concerns the *difference* between Can and $Does$, it lies in the fact that Can has no statistical or normative power, but instead represents a *possibility*, hence forming a more elaborate basis for social planning and anticipation. Therefore, we are also not restricted to deriving skill information from observation solely, we can model it through communicated skills (preassuming trust!) or by combining observed information to construct assumptions about skills not yet observed (e.g., in a monotonic model, we might use a rule like

$$Can(R, S_1, S', \phi_1, cond_1) \wedge Can(R, S', S_2, \phi_2, cond_2) \rightarrow Can(R, S_1, S_2, \phi_1 \wedge \phi_2, cond_1 \wedge cond_2)$$

to infer complex skills).

4.1.2 Intentional attributes

The intentional level of role models captures knowledge about the goals, beliefs and preferences of a role. Generally, these cannot be observed directly, and although assumptions about the rationality of the agent or group that is fulfilling the role can be combined with action observation to infer knowledge at the intentional level, such knowledge will always be approximate and uncertain. Here, communication may provide additional information, but the basic problem of facing the other as a "black box" cannot be done away with completely, since the other might be lying or ignorant about its own intentions.

Without intending to dive into the epistemic subtleties of reasoning about knowledge and intentions, we sketch modelling primitives that may be useful even in the absence of certainty. Their usefulness lies in their explanatory power: knowing what the other "believes" and "wants" can be used together with what the agent/group "can" do to infer what it *will* do.

We should stress the fact that the "mentalistic" categories used here are of subordinate importance to our framework which is trying to avoid the "psychological trap" in talking about interaction (it does so by focusing around "social action" rather than on the internal states of agents). In fact, they would be negligible were it not for communication, which obviously enables agents to inform others about their internal state.

Values. Presuming that roles fulfil the minimal assumptions laid out in section 2.1, agents fulfiling them must exhibit some preferences regarding their state of belief. Although a statement of the form

$$Prefers(R, S_1, S_2)$$

stating that R prefers situation S_1 over situation S_2 provides less information than knowledge about whether R is actually pursuing S_1, it may be valuable as some sort of proto-intentional information, since the way R forms its intentions is based on its preferences, at least if we assume rationality in making decisions.

Consider, for example, a world that is in state S_2. If agent a knows that $Prefers(R, S_1, S2)$ and believes b to fulfil role R, it may try to convince b to help a in bringing about S_1, and this may be of crucial importance for both, if they both cannot achieve S_1 alone.

A further possibility is to make *quantitative* statements about preferences rather than merely ordinal *qualitative* ones. For reasons of simplicity, we will refrain from inferring such quantitative information e.g. in the form of utility functions for the time being. Note, however, that this bears severe implications for the scope of our approach, because, to name just one limitation, we are not capable of distinguishing between risk-neutral, risk-averse and risk-taking agent models.

Tasks, Goals and Intentions. If goals are defined as states of affairs more desirable than others (as in Markov Decision Processes, in the state-oriented domains of [19] or in traditional AI search and planning [14]), they can be reduced to the above *Prefers* predicate, since such a view requires no more than a valuation over situations. However, if having a goal is already affecting the agent's behaviour (as in the case of actively pursued goals, or, even mor specific, in the case of tasks in progress), *intentions* are involved, i.e. decisions which imply a commitment to behave in a certain way. For these, we might suggest the following predicate:

$$Intends(R, S_1, S_2, \phi)$$

We use this syntax to denote that in situation S_1, R actively pursues the achievement of S_2 by carrying out ϕ (which, again includes at least one of R's action symbols but none of other roles). This view implies that

1. The intention is triggered by some state of the world which makes the agent want to change it. There is a *motivation* for doing something and this consists of a *difference* in the desirability of situations.
2. The agent has some *plan* of how to do things that it has committed itself to. We make no assumptions about whether this plan actually will bring about the desired changes, but we know that the agent will use it. Also, this view implies no deontic social semantics of commiting oneself to something towards someone, which may exist if commitments are *communicated.*
3. There is no such thing as a *joint intention* (as in [6]) in which a group's intention cannot be seen as that of a single actor. Being a huge simplification, we believe that it is consistent with our view of interaction, since joint intentions are the product of interactions *within* a group and hence detached from a frame describing an interaction *towards* that group. This means that if a is interacting with a group G of agents, it will be treating that group as a single agent.

It is arguable whether such an intention can be asserted without R believing that the plan will lead to the desired state, i.e. usually we would assume that

$$Intends(R, S_1, S_2, \phi) \rightarrow Believes(R, (Now(S_1) \wedge \phi \rightarrow Eventually(S_2)))$$

and, likewise, intentions imply many other mentalistic assumptions which we will not discuss here (see, for example [18] for a more detailled account). For our purposes, it shall suffice to know about a role how it forms its commitments under particular conditions. Unlike the *Does*-predicate, intentions enable us to represent genuinely *normative* models of behaviour thus escaping the trap of building expectations solely through observation: in a multiagent planning initiative of a single agent, it might for example create an imaginary role of a peer who "plays her part" in executing the plan by using intentions[6].

Beliefs. The very general category of belief is the final mental primitive that we use in modeling others by using a simple predicate

$$Believes(R, S)$$

to denote that R believes S (any logical statement). As in the case of joint intentions, joint beliefs are not further decomposed, since whole groups are seen as a single entity with which to interact.

With respect to the modeling layers put forward in figure 1, information about beliefs is located at the highest level of model complexity, since it represents knowledge about

[6] *Prefers* and *Can* can also be used to define normative expectations, but not about anticipated *behaviour*, only about a stance or an ability.

the internal states of the agent/agents that are being modelled.

Most readers will be reminded of the theory of BDI agents [13] by our account of modelling role properties at an intentional level. Indeed, what we present here is a simplified version of BDI models (simplified with respect to the fact that it assumes fixed preferences in the place of "desires").

4.1.3 Third-order (social) attributes

Having described aspects of modeling the individual properties of an agent or group that is enacting some role, we now turn to the *social* dimension of roles, recalling the equation put forward at the beginning of section 4.1 which states that a role is not only defined by its expected behaviour but also by its relationship to other roles. To see why this is the case, we should ask why any more than expectations about the individual would be useful for someone interacting with actors who fulfil the role in question. Would it not be sufficient to have some model of what these actors can/will actually *do* and *think*?

Admittedly, from a reductionist point of view, it would. For example, a statement of the form

$$Commands(R_1, R_2)$$

saying that R_2 will intend whatever R_1 requests from her can be transformed into something like

$$\forall X \in A_{R_2}.Requests(R_1, R_2, X) \rightarrow Intends(R_2, \text{true}, \text{true}, X) \quad .$$

However, there are two good reasons to treat information about the relationships between roles as first-order concepts:

1. As stated in section 3.2, relationships may provide missing information that goes beyond the trajectories that are defined as an intergral part of the frame: they reify interdependences that are more persistent in a society than the trajectories of interactions enacted in a single frame. To return to the above example, since the *Commands*-statement only refers to actors and not to situations or particular interactions, it may hold throughout various frames that involve R_1 and R_2.
2. They encapsulate information about the relationships between *roles* and are thus located at a different level of abstraction: the specifics of the definition of a role (in terms of behavioural and intentional attributes) are no more affected except as to their dependence on other roles. In the above example, it would be counter-intuitive to say that the lengthier (second) formula *defines* R_2 or R_1 as such, because it always depends on (at least) a second role and thus contradicts the intuition of the vocabulary for role properties we have defined so far.

Here, we are going to discuss two types of social attributes: rights, duties and responsibilities towards others on the one hand, and social position within social structures on the other.

Rights, duties, responsibilities. The rights, obligations and responsibilities that an agent has towards others form a basic category within social worlds, and, more specifically, an inherently normative category, i.e. they prescribe what the agent is supposed to do and not to do. Keeping in mind that we have based our notion of frames around a set of *trajectories* that define the possibilities of action for participating role-representants, we are faced with a dilemma: if the "trajectories" slot says what people ought to do and these norms are necessarily constructed in terms of the interacting "roles", than what kind of additional information could be provided by associating "extra" norms with the roles themselves?

One possible answer would be that we want to factor out norms from frames and associate them with roles that are valid throughout several frames. But in that case, frame inheritance (cf. 4.3.3) would do the same job, because sub-frames inherit trajectories from super-frames. Another possibility would be to include such role-bound norms to make a distinction between them and frame-bound trajectories, but in that case, they would be detached from the interaction the frame is about, and this contradicts our intuition that all information about action must be directly related to interaction.

Both these views obviously allude to a redundancy between role-bound and frame-bound norms. In spite of this redundancy, we suggest allowing both types in our data structure, for the reason that role-bound norms may reassure agents in the process of framing even if they are not relevant in a particular enactment of the frame. As an example, even though a specific classroom-lesson frame may not directly include expectations about whether a teacher is going to beat the students, it is very reassuring for the pupils to know that a teacher is generally not allowed to beat students. Hence, the classroom-lesson frame becomes less risky and students are more prone to participate in it. Also, such information provides continuous reassurement about role instantiation: as long as the teacher is not violent, we are convinced that she may still be considered to be a "teacher" (otherwise we might also switch to thinking that she is a tyrant).

Since the vocabulary we use to formalise such socially normative aspects of role models is identical to that used for frame-norms, we postpone its definition to section 4.2.

Social position, status In the social sciences, the terms "social position" and "status" most often define the (genuine or attributed) position of an individual in a system of social relationships (segmentation, aggregation, etc.) which somehow structure the social system.

In our framework, since we reason only about roles and not individuals, we can describe the social position of a role by its relationships to other roles (for which a basic terminology will be introduced in the following section).

4.1.4 Relationships

The categories of relationships we account for are *dependence*, *acquaintance*, *aggregation* and *segmentation*. Although certainly not complete, this list includes the most relevant categories to describe how roles relate to each other.

Relationships of dependence are used to express control, power and similar phenomena between roles. In our interpretation, dependencies exist with respect to (possibly complex) actions, not simple states of affairs, because they should directly relate to an agent's capabilities (conditions/situations do not directly define the options an agent has, and we view dependencies only as dependencies that are *explicit* in the social model and require no further inference to be derived). Thus, if a relies on b to open the door for a, it does not depend on b to achieve the condition "door open", but on b to open the door so that "a may walk though it". The predicate we will use for these dependencies will be

$$Depends(R_1, R_2, \phi_1, \phi_2)$$

to indicate that R_1 depends on action ϕ_2 to be performed (or not performed) by R_2 in order to execute ϕ_1 (henceforth, indexing action formulae ϕ with the indices of roles shall imply that the formula includes at least one of the action symbols of that role and no action symbol pertaining to other any role(s)). As before, we neglect cases of purely cognitive phenomena (e.g. "mind control"); they are deemed irrelevant, if they don't bear consequences on action.

Acquaintance is a relationship we have to deal with in order to formalise that roles have some information about each other. In the most general sense (by interpreting "information" as "making a difference"), this is the case whenever role R_1 *reacts* to something R_2 *does*. In a certain sense, acquaintance is the "minimal" variant of interaction – it means that the "other" makes a difference. In contrast to dependence, though, the other's action ϕ_2 is only a *sufficient*, not *necessary* condition to react by performing ϕ_1. A predicate of the form

$$Acquainted(R_1, R_2, \phi_1, \phi_2)$$

will be used to express this relationship, with role identifiers and action formulae as before.

Relationships of aggregation are necessary to group n-level roles $R_1, \ldots, R_n$ to an $n+1$-level role T (where $n = 1$ is the basic case of single-agent roles) and can be expressed by using a predicate of the form

$$Member(R_i, T)$$

which is strict and binary, i.e. there is no "degree" of membership or "must"/"may" dichotomy, unless otherwise modeled by the language we use. Membership need not be "formal" in the sociological sense, it may be transitory or even hold only for a single interaction frame. Further, membership relations may overlap, i.e. a role R may pertain to several aggregates.

As the opposite of aggregation, segmentation is needed to mark differences between roles. However, these ultimately result in differences in expected behaviour, ability, mental state etc. which are explicitly modeled when introducing the roles. Considering segmentation that is due to *labeling* irrespective of explicit differences between roles, a phenomenon which is very common to human societies, we assume the stance that this is due to *attributing* a certain role to someone (the "criminal", the "punk" etc.), but that the specifics of the role itself are still explicit expectations towards it. Hence, we prefer to derive relationships of segmentation from the aforementioned categories rather than to ascribe a status of first-class objects to them.

This ends our description of roles as abstractions from individuals. We believe that our formalisation provides an expressive language for defining roles and their relationships towards each other by exploiting the basic categories of *behaviour expectation* and *interdependence* (basically, all other concepts are only specialisations of these).

We round up this section with a brief note on *role repositories* the construction of which we consider helpful. The basic and very simple idea behind them is that, since roles may be used beyond particular frames a role (and relationship) database should be built up by agents to generate and maintain role knowledge. Such a data structure that allows grouping, dependency structuring and labeling with frame identifiers (to restrict the validity of statements to certain frames) is sketched in figure 2.

4.2 Abstracting from actions: trajectories

Trajectories are, apart from roles, the second central building block that constitutes a frame. As mentioned, they encapsulate knowledge about trajectories of action by defining what must, may and must not be done when enacting a frame. In deriving trajectories from observation, a basic *duality* becomes obvious: on the one hand, whatever agents do is in some sense normative since it establishes some kind of "practice". On the other hand,

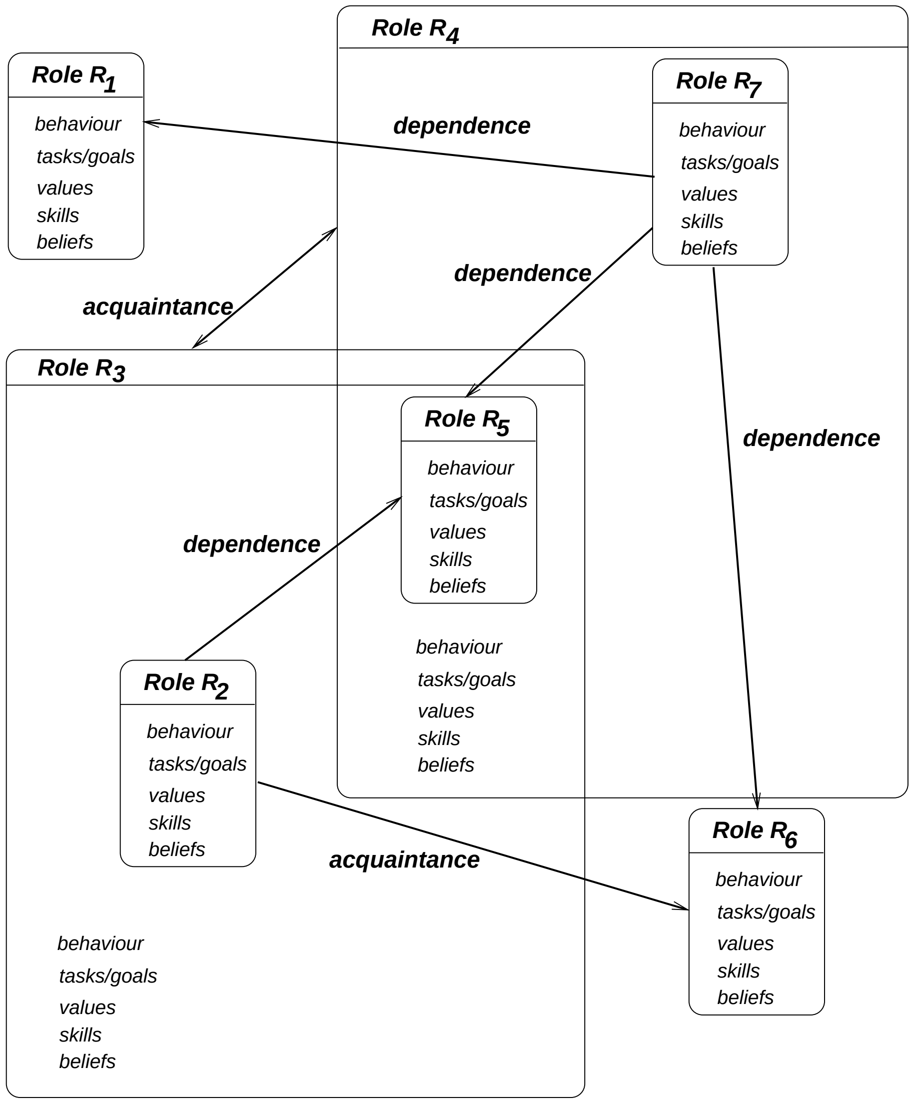

Figure 2: (Sketch of) A possible role repository data structure: roles are depicted as boxes, aggregation and dependence/acquaintance is shown by grouping roles to super-roles and by using labeled arcs between roles, respectively. All elements (attributes, roles, relationships) might be labeled with the names of frames to restrict them to these frames (this possibility is omitted here for the sake of readability).

agents may be acting in contradiction with existing norms, and the seemingly paradox effect may occur that they may be *expected* to do so (a criminal is expected to commit criminal actions which violate societal norms and values). To relieve this problem, we can refer to the logic of symbolic interactionism: there can be no "value" or "norm" unless it is enacted by interactants with some regularity. If a criminal commits a crime (which he is expected to do), society sanctions him, in general, or at least attempts to. That sanction is part of a norm between state and citizen, i.e. a reaction to deviance. But in the case of a state-criminal interaction, it is a norm and not a sanction, because the criminal is not deviating from his expected behaviour. Although this view may not appeal to sociologists in general, we will adopt it here to obtain a value-free definition of norms which is not only intuitive because of its pragmatic character but also suits well a notion of "unbiased" interaction among artificial agents.

We introduce the following predicates to formalise norms:

$$Must(R, S, \phi) \tag{1}$$

$$May(R, S, \phi) \tag{2}$$

These denote obligation and permission (as common to deontic logics, sometimes also "must" and "may" are replaced by modal operators in possible worlds semantics), thereby spanning "spheres of commitment" [16]. Quite deliberately, the condition S under which such a commitment arises need not include action symbols and hence it does not necessarily depend on (others') actions.

As remarked before, we do not intend these primitives to be used for general norms only, but also for defining very specific trajectories in interaction frames, which may range from simple protocols like a handshake to highly elaborate workflows in complex domains. In either case, the important thing to note is that, if assumed by an agent in a frame it is activating, they represent the rights, duties and responsibilities of some role with respect to a frame (if they are globally valid with respect to a role, they apply as discussed in 4.1.2) and thus bridge the gap between expectation and reality by inducing commitments on the interactants. A schematic view of trajectories, very similar to protocol diagrams in AgentUML is provided in figure 3.

In the following section, we turn to attributes that describe *meta-frame* knowledge. While the definition of what constitutes a frame is complete whenever the respective rules and norms are accounted for, these meta-attributes provide the basis for activation and interpretation of roles, and also for linking several frames with each other semantically or historically.

4.3 Meta-level knowledge

Introducing roles and norms in a frame data structure yields a description of interactants and their actions with respect to an expected pattern of behaviour, which is useless unless (1) conditions for its use are specified and (2) it is put in the context of other existing patterns. To cater for (1), we introduce activation "contexts" and "meta-frame beliefs" in this section, and (2) is accounted for by using meta-frame "histories" and "links". Thereby, contexts and beliefs provide information about "how and when to frame" while histories and relationships connect frames with each other.

4.3.1 Context

In short, contextual rules define when frames should be activated by their user. We account for five possible contextual categories which can be expressed by using the following

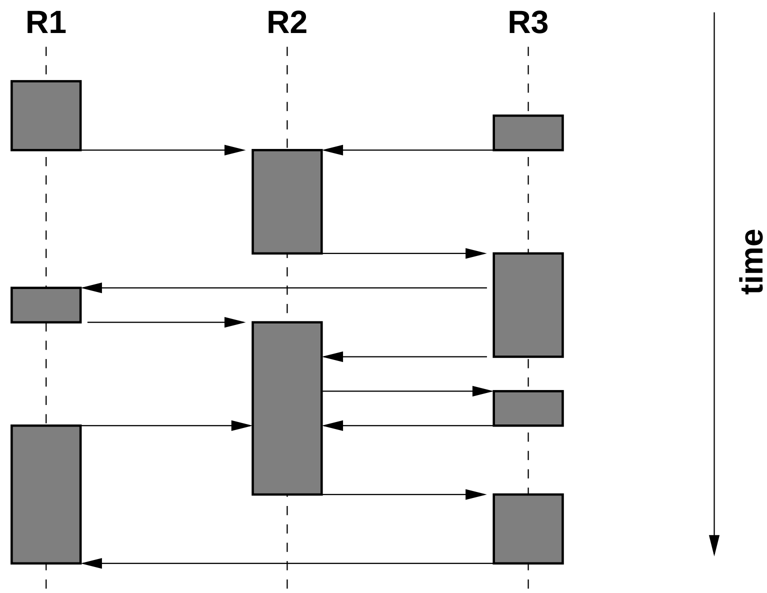

Figure 3: Graphical trajectory schema. Lines of actions are depicted along the vertical timeline, with actors (represented by roles) executing their parts in parallel, and horizontal arrows denoting reactions of one role to some other role's action. Semantically richer representations can be used, e.g. by labeling arrows with message contents, using logical connectives to include several paths of execution, etc.

predicates:

$$PreCond(S)$$
$$PostCond(S)$$
$$Maintain(S)$$
$$ActCond(S)$$
$$DeactCond(S)$$

$PreCond$ abd $PostCond$ are used to denote necessary pre- and postconditions S, which must hold/will become true before/after the trajectory is put into practice. Likewise, $Maintain$ describes what should hold while the frame is active, i.e. whenever $\neg S$ is the case, the frame is jeopardised.

Unlike these two predicates that only loosely "associate" certain situations with the frame, $ActCond$ and $DeactCond$ provide clear-cut *conditions* for activation and de-activation. Their choice is very crucial, since they should "switch" frames "on" and "off" – in the simplest case, there would be exactly one frame that is activated in any situation S.

Using these categories of conditions, we obtain an outlook on trajectories as "embedded" within activation/decativation rules and linked to pre-, post-, and sustainment conditions for successful execution, as is depicted in figure 4.

It should be noted that activation/de-activation conditions need not be complementary; in fact, in most cases they will more or less resemble the pre- and postconditions of a plan, and the same holds for maintenance statements. Providing extra categories of conditions for activation and de-activation, however, detaches some of the conditions as "decision-triggers" from other conditions, so that the private stance of the agent towards the frame can be treated separately.

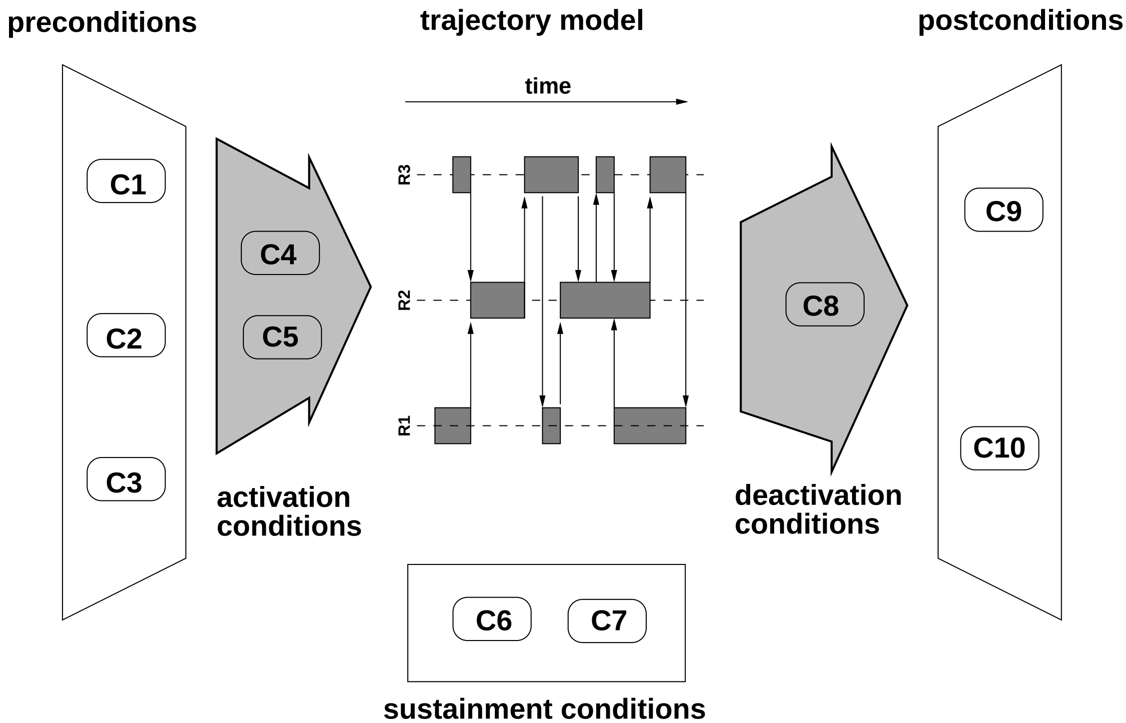

Figure 4: Graphical context model. The frame trajectory is embedded in activation and deactivation conditions (shown as shaded arrows) and surrounded by an environment of pre-, post- and sustainment conditions (in boxes).

Also, most likely, implementations will use quantitative versions of these predicates to express uncertainty (e.g. enriched by fuzzy values) unless frames are used very rigidly as fixed protocols, since there is nothing that should render the agent certain about when to activate what. Therefore, the suggested predicates should only be regarded as prototypical.

4.3.2 Beliefs

Although beliefs about the world form a very important part of interactions, we choose to separate them from actual role, trajectory and contextual knowledge. Mainly, this is a consequence of our symbolic interactionist outlook on social action, according to which meaningful interaction occurs, whenever lines of *action* fit together – the cognitive phenomena associated with social action may be the causes and effects of social action, but they are not the focus of reasoning about action. Therefore, interaction may be managed very well even without interactants knowing anything about each other's beliefs; moreover, as long as the "allignment" of actions works out for the interactants, they need not even *bother* about each other's mental states.

However, even though – strictly speaking – information about beliefs does not form part of the interaction process, it can be very useful as information about epistemic conditions for enacting the frame. For example, a shared interpretation of communication primitives may not be necessary for enacting a frame that involves communication (as long as the right communicative actions are taken), it surely supports confidence that the frame can be carried out properly. Likewise, knowledge about the world may be important though not essential: it is enough for "alter" not to plunge her hand in the boiling water

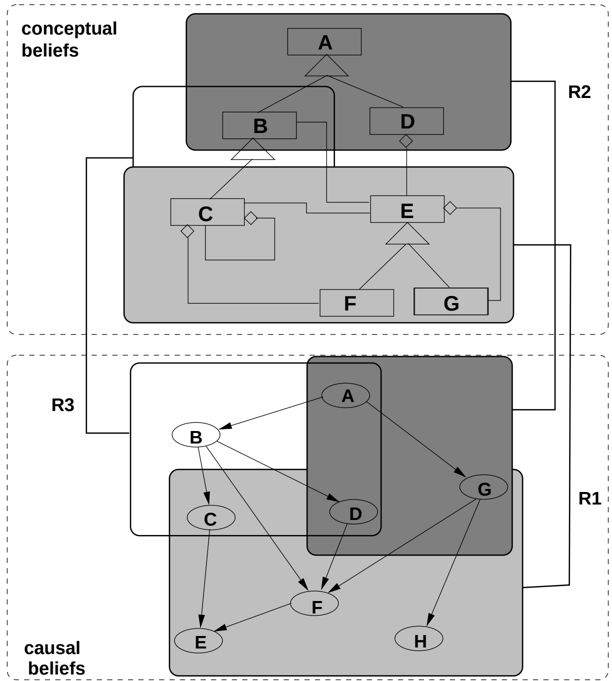

Figure 5: Belief model, with roles' beliefs as (shaded) subsets of taxonomies (we use an UML-like notation for simplicity) and belief (here: Bayesian) networks.

for a "cooking together frame" not to be jeopardised, but the situation would be safer if "ego" could assume that "alter" actually *knows* that the water is hot and that hot water causes injury. As a third category of useful epistemic conditions, consider "recursive" frame knowledge: returning to the "classroom lesson frame", it is easy to see that it is enough for the children to simply leave the classroom when the lesson is over, but if the teacher can rely on them knowing that she will de-activate the "lesson frame" at the ring of the bell (and that no student will expect her to go on with the lesson after that), her expectations about the interaction will be less contingent.

We will introduce no further formal elements to state such beliefs, since we believe that the primitives previously introduced should (together with domain-specific elements) be sufficient to express most conceivable epistemic conditions. Figure 5 merely sketches a very general notion of belief structures, one in which the knowledge of participating roles is defined by subsets of *conceptual knowledge* (ontologies, taxonomies, semantic networks) and/or *causal knowledge* ((probabilistic) rules, production systems, etc.).

4.3.3 Frame histories

Frame histories are a very simple means of tracking the development of a particular frame by recording a sequence of modification operations $op_1, \ldots, op_n$ to an original (empty) frame, from which the current frame evolved. Tracking changes to a frame may be useful for various reasons:

1. Conflict/Uncertainty: if an enacted frame does no longer suit the agent's purposes or if the agent is not certain about whether the frame will work, it may need to return to previous (perhaps better established) versions of it.

2. Learning: in search of "better" frames, agents will modify frames constantly; if these

modifications prove inefficient, there should be a possibility to go back to previous versions.

3. Recording experience & Evaluation: Tracking the changes a frame undergoes enables the agent to store and review its past use of the frame.

Located at a meta-frame level, these histories can be formalised by referring to the current frame as `this` (and possibly to other frames $F_1, \ldots, F_n$, e.g. in case `this` is the product of parallel composition of two frames F_1 and F_2) and noting operations as modifications to the data formalised in the previous sections. Possible operators would include

$$\begin{gathered}
AddNorm(\ldots) \\
RemoveNorm(\ldots) \\
AddRole(R) \\
RemoveRole(R) \\
ModifyRole(R, Replace(Skill1, Skill2)) \\
AddBelief(\ldots) \\
RemoveBelief(\ldots) \\
AddContextRule(DeactRule(\ldots)) \\
SeqCompose(\texttt{this}, F_5) \\
\vdots
\end{gathered}$$

although this list is certainly not complete.

4.3.4 Frame relationships

There are two categories of relationships between frames that are worth considering in the "links" slot: firstly, purely *syntactic* relationships (part-of-relationships, subsumption/inheritance, aggregation) that can be useful when managing complex *frame repositories*. Secondly, *semantic* relationships that make existing relationships between the attributes of the respective frames. Examples for these include

$Alternative(F_1, F_2)$	holds when the two frames apply in similar conditions, use similar roles and similar trajectories
$Contradicts(F_1, F_2)$	valid if the two frames cannot be enacted between with intersecting sets of interactants in parallel
$IsMoreSpecific(F_1, F_2)$	F_1 makes "more" assumptions about roles, relationships, trajectories and beliefs (not the same as "technical" inheritance)
$Variant(F_1, F_2, Diff)$	F_2 differs from F_1 with respect to $Diff$ (can be used to express that frames share roles, trajectories or beliefs, but also to express more subtle differences)
$\vdots$	

and, surely, many more are conceivable, which we cannot account for at this point in time.

This concludes the section on "common" frame attributes, which are sufficient to define the purely *social* side of interaction frames. However, the process of "framing" is

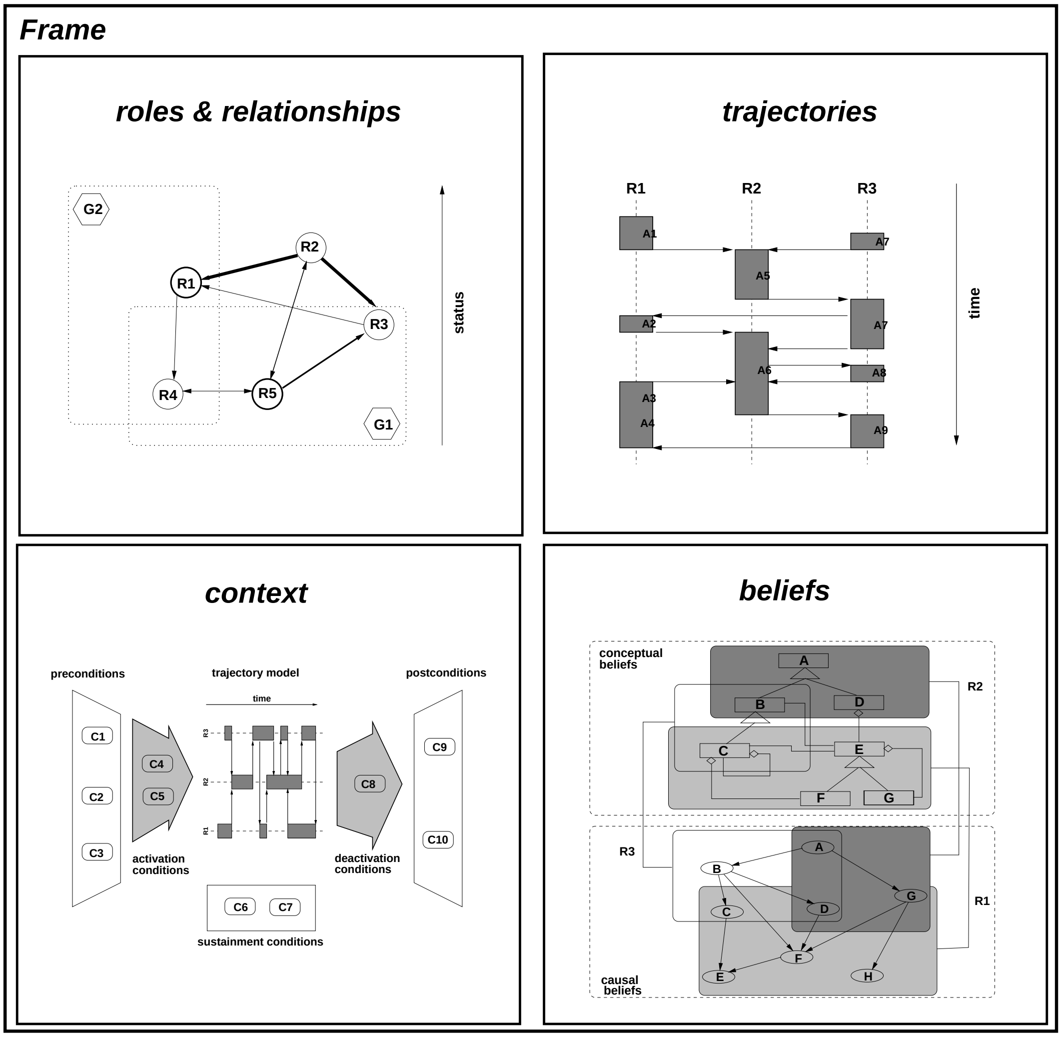

Figure 6: Frame model with roles and relationships jointly represented in the top left diagram, trajectory model sketched top right and context and belief schemas as previously introduced. The links and history attributes are omitted as well as the private attributes.

an activity that must be left up to the individual, and, hence, the individual will need to supplement its knowledge about an interaction pattern with information about its own *stance* towards the frame in the current situation. This process, which essentially can be reduced to asking is explained in detail in the forthcoming section. In discussing the process of framing, we will often use a simplified graphical represenation of frame that omits the "links" and "history" sections of the common attributes and also all of the private attributes for the sake of readability. A possible notation is introduced in figure 6.

5 Framing

Framing is a very complex process which includes (1) tracking the enactment of activated frames, (2) choosing whether to retain the current frame or to change frame when appropriate, (3) modifying frame knowledge with experience and (4) relating these three activities to one's private goals in order to make them part of individually rational decision-making and behaviour.

This section is divided in two parts: in the first part, we present a rough outline of the framing process as we see it in order to provide an overview of the interplay between the activities involved. In the second, we take a closer look at components that are necessary to manage the process and propose an abstract agent architecture to cope with it.

5.1 The process of framing

Figure 7 shows a simplified model of the framing process as a mental function of the framing agent that mediates between perception and action. The data structures involved are

- the *current frame* (denoting the unique interaction frame currently activated),
- the *difference model* (derived from a comparison between perceived situation and current frame),
- the *adjusted frame* (containing the current "hypothesis" whenever alternatives for the current frame are sought for)
- and the *frame repository* which contains a (suitably organised) frame database as a "hypothesis space".

As the agent continually perceives its environment, these data structures provide the following functions with relevant data in order to make framing decisions:

1. *Matching:* The current situation is matched against the current frame (which is supposed to describe it, if accurate) in order to generate a difference model that identifies which aspects of the interaction comply with/deviate from the current frame.
2. *Reaction:* The difference model is used to decide whether to retain or abandon the current frame.
   (a) *Derive commitments:* If the agent chooses to adhere to the current frame, then that frame is used in a prescriptive way in order to infer commitments for the agent that arise from the frame. Also, the frame repository is updated to ensure consistency with new experience.
   (b) *Adjust current frame:* If the agent chooses to deviate, the frame has to be adjusted in order to match the situation. Consequently, "adjusted frame" hypotheses have to be formed (using the frame repository) and to be forwarded to the *matching* function ("trial instantiation"), until a suitable frame is found.

To sum up, the process ensures that the current frame is consistent with the agent's perception by spawing a modification process whenever the degree of "deviance" is not acceptable.

As far as the "private attributes" of frames that were introduced in section 3 are concerned, they relate to the framing process in the following way: in the matching process, they are filled with information that is obtained by matching the perceived situation with the frame currently activated. The "evaluation" slot, however, is not modified until the *reaction* function, in which private goals enter the scene.

Whether or not all of the slots are also supplemented with "reasons" for particular status assumptions as suggested in table 2 is left to the designer. The same holds for the actual representations (fuzzy, probabilistic etc.) of status knowledge that will be used.

This simple view of the framing process raises a number of questions and possible points of criticism:

- *What if no matching frame can be found?*
  It has to be ensured that there is always a unique "most general frame" that matches *any* situation by representing a mental state in which the agent knows nothing about the current interaction. In fact, in a purely adaptive setting, this would be the only frame known to the agent at the beginning.

- *Isn't comply/deviate too simple a dichotomy?*
  Indeed, we do not mean complete conformity or utter deviance by using these terms. However, we do have to draw a line between when to maintain and when to abandon a frame. For example, even if a frame is merely "completed" (i.e. the interaction ends as expected), a new frame has to be sought for to continue. Or, if some peer does no more comply with a role model it was assigned, we have to choose whether to alter the role model (and with it, the frame) or whether to ignore the deviant behaviour. Or again, as a third example, if one is deviating against a sanctioned norm, she has to decide whether the consequences are simply "endured" sanctions, or whether some other frame is available that might help in avoiding them.

- *If frames are updated with interaction experience, do they not grow overtly specific?*
  It is an intended feature of the suggested procedure that frames show an inherent tendency towards specificity. This means that the more they are confirmed by being retained, the more specific will the properties of their roles, trajectories, contexts and beliefs grow. This is mainly due to the fact that the properties of the current situation continuously feed into the current frame as additional information. Thus, contradiction with future situations will always be necessary to generalise from the specifics, and this is precisely the behaviour we want to achieve – eventually, the frame that is constantly going "back and forth" between generality and specificity is hoped to converge to an optimal level in describing the concerned interaction.

- *More generally, isn't there something like a "frame problem" the agent has to face?*
  The answer to this question is: yes, positively. Each new facet of an ongoing interaction that is being experienced might be dealt with in different ways: it might lead to an extension or restriction of the current frame attributes, to a search for an adequate alternative frame or to the creation of a new frame; it might necessitate a merger of frames, or a "split", etc. Thus, the agent is constantly faced with the question of how to group its experiences together to "interactions" in a reasonable way, and finding answers to this question will be a key issue in our future research.

In the next section, which represents the "essence" of our considerations so far, we combine the concepts we have developed to an integrated prototypical agent architecture that is based on frames and framing.

5.2 A "framing" agent architecture

The agent architecture we propose is, strictly speaking, a *social cognition* architecture, i.e. it focuses on the mental activities necessary to manage interactions effectively, or, in other words, to *coordinate* [7]. One of the implications of emphasizing this aspect is that we completely ignore others, in particular: acting and sensing, prioritizing goals, planning and monitoring plan execution, as long as these processes are non-social. This enables us to set up an entire reasoning and deliberating machinery around the notion of interaction, thus ensuring (1) that we conform with the social theory we have used as a

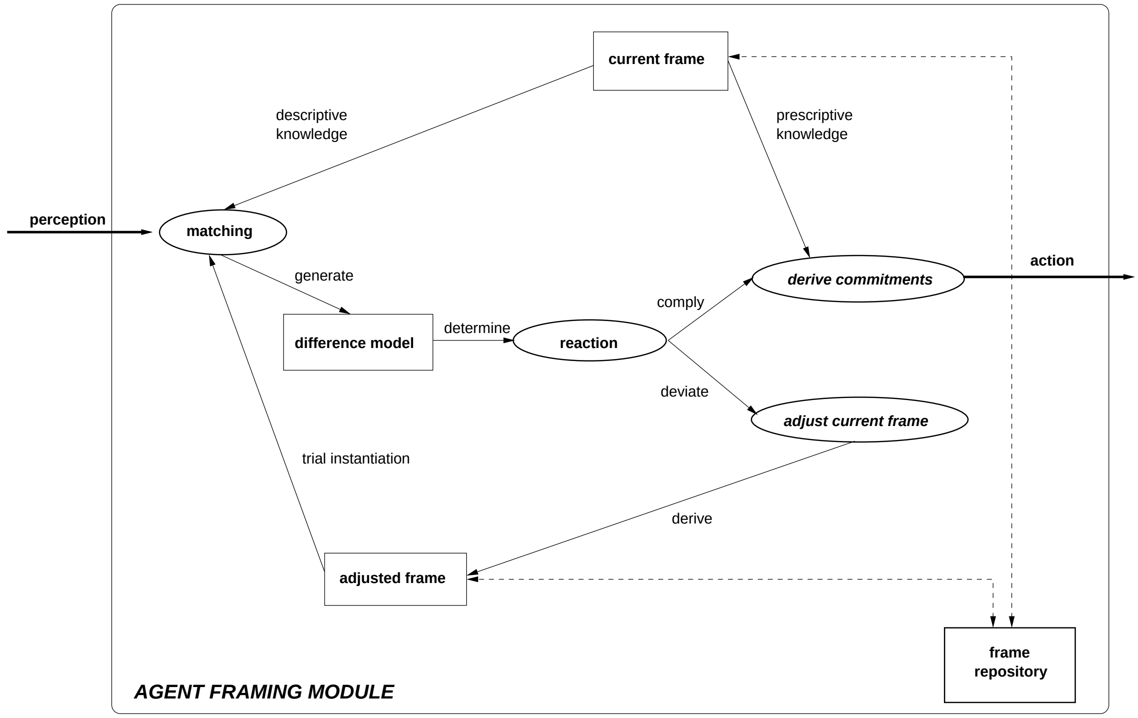

Figure 7: Simplified overview of the framing process. Data containers (current frame, adjusted frame, difference model and frame repository) are shown as boxes, functions in ovals.

starting point (2) and that we model agents as neither solitary, nor socially predetermined but as "socially oriented" yet "selfishly autonomous" entities.

The architecture, depicted in full detail in figure 8, refines the framing process described in the previous section by introducing several functional components and by making the processes of matching, reaction, adjustment and commitment more concrete.

To begin with, in contrast to the picture laid out in the previous section, percepts have to be interpreted in terms of a "perceived frame" (that is constructed independently from the "current frame" data structure). Interpreting recent percepts as elements of roles, trajectories, contexts and beliefs is necessary, because a simple percept cannot be directly compared to the currently activated frame, which is a fairly complex data structure that preassumes the identification of action and communication sequences, their classification, generalisation etc. Hence, it is not *perception* as such that is compared to the current frame to obtain a difference model, but a frame-based model if the *interaction situation* constructed from information that must be derived from recent percepts. To this end, the perceived frame has to be updated with new percepts and provides the "social world model" for the situation interpretation module.

Then, again refining the previous outline of framing, the frame matching module compares the perceived frame to the currently activated frame. Here, the two frames play the roles of *descriptive* ("the way interaction is") and *normative* ("they way interaction should be") models of interaction. The difference model that results from the matching module's computation basically incorporates a classification of roles, beliefs, contexts and trajectories into compliant and deviant "parts".

The frame assessment module, previously presented as a simple decision function, actually comprises three functions that it performs on the difference model and whose results it has to combine to make a framing decision:

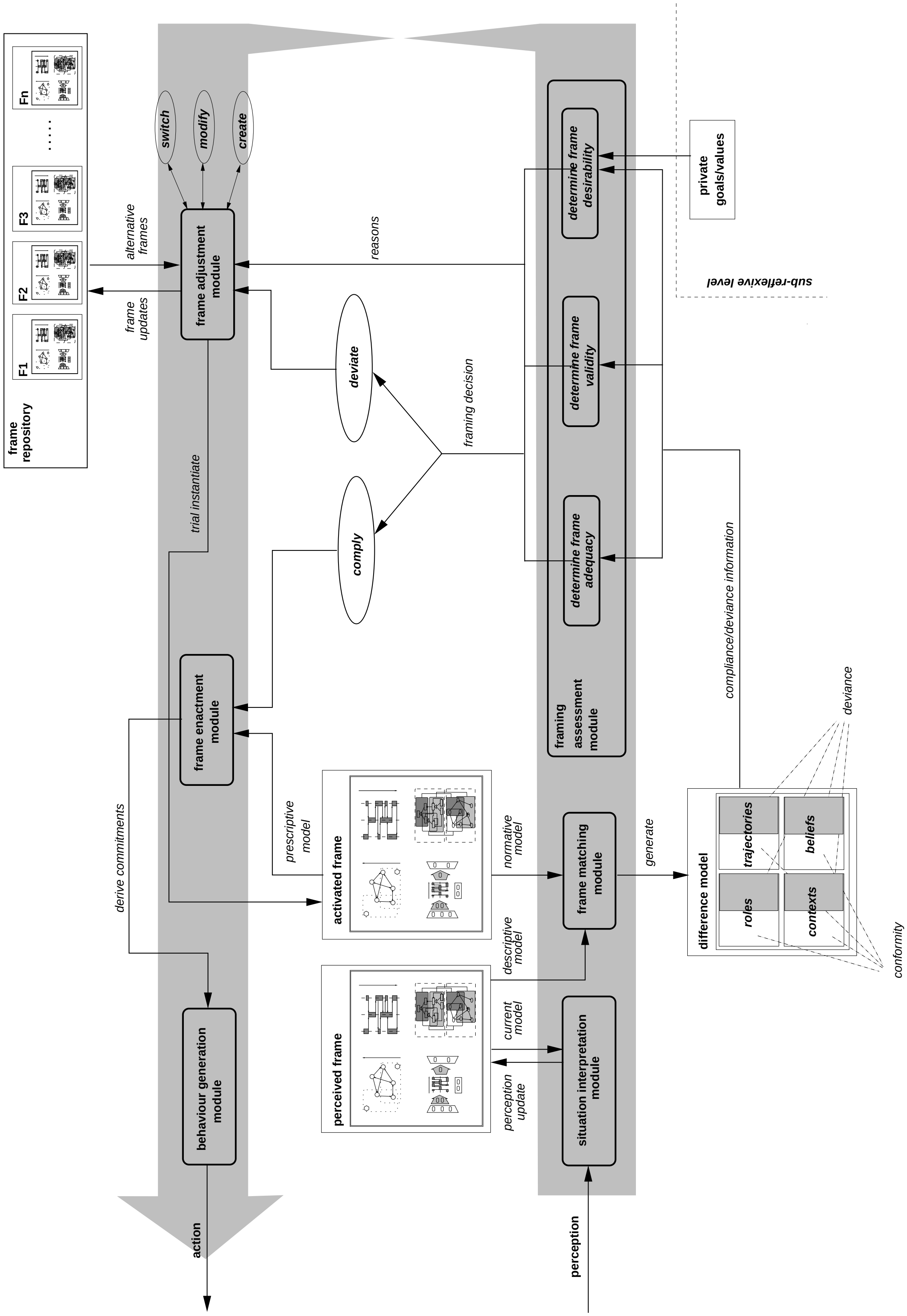

Figure 8: Detailed view of the framing-based agent architecture. The main line of reasoning between perception and action (shown as a shaded arrow) captures both the sub-processes involved and the temporal order in which they occur. Frames are shown using the representations introduced in figure 6, and links between them (as well as histories) are omitted in the "frame repository" for the sake of simplicity.

1. *determine frame adequacy:* this function assesses the validity of contexts, i.e. it determines the agent's reaction to the conditions that are fulfilled by the current situation and those who are not; this is the basic "filter pass" that is used to decide whether to retain or to abandon a frame – if too many conditions are violated, the current frame is judged inadequate;
2. *determine frame validity:* as a second "filter pass", this function determines appropriate reactions to the "degree of enactment" that the perceived frame represents compared to the activated frame; obviously, if the interactants do not match the assumed roles and relationships, if they do not adhere to the expected trajectories or if they do not appear to share the necessary beliefs, frame adjustment has to be considered;
3. *determine frame desirability:* the final "test" the difference model has to pass consists of evaluating it against the private goals (that are considered external to the framing module) of the agent, since, even if the perceived frame perfectly matches the activated frame, it has to be *useful* in order to be kept up by the agent.

This module makes the basic decision about whether to comply with the current frame or whether to deviate and is therefore one of the most crucial component of the whole architecture (in a certain sense, it controls the agent's *exploration/exploitation* behaviour).

If the agent choses to conform to the frame, the frame enactment module derives commitments from it – after all, it is the very use of a frame to guide the agent's behaviour as a *prescriptive* model of action and actors (behaviour generation module). The choice of the frame of course also feeds back to that very frame, since the decision underpins (by adding the decision itself and possibly also the reasons for it to the frame data) the normative character of that frame.

In case adequacy and/or validity and/or desirability assessment produce a re-framing decision, two basic phenomena can be distinguished:

1. Framing uncertainty: if the frame is inadequate or invalid, the agent is not sure about whether the frame still adequately models the interactions it engages in.
2. Framing contempt: if the frame is no more desirable, the agent sees itself involved in an individually unacceptable (at least undesirable) interaction process.

The effect common to both phenomena is, initially, the *inhibition of action*: the agent cannot simply continue enacting the frame, and is thus forced to reconsider its frame models. In our simplified AI view of interactionist theory, we can simply postulate that "there is no un-framed situation", i.e. before interaction can be continued after inhibition, the agent *must* activate some frame, if only the "empty" frame that represents the riskiest and most uncertain of interaction models.

Accordingly, we suggest the use of a *trial instantiation-matchmaking-assessment* cycle: in each iteration, the frame repository is used together with reasons for deviating provided by the frame assessment module in order to generate adjusted frames (frame adjustment module) which are then forwarded in a "trial instantiation" fashion to the matching component and evaluated by the assessment module. This iteration is repeated until a suitable frame has been found that the agent chooses to comply with.

Looking very simple, this control loop has to be engineered very carefully in practice: firstly, since there is a vast variety of frame adjustment options (modify, switch, create new frame, merge, split etc.) the search must be guided by suitable heuristics. Secondly, the assessment module must be forgiving and strict at the same time and compromise the behavioural tendencies of others with the agent's private goals. Thirdly, the inhibition induced by searching for alternatives must not take too long, because "not acting" creates additional uncertainty on the side of peers, and this mutual irritation might even culminate to a complete "interaction breakdown" in the society.

The proposed agent architecture completes our account of interaction frames for artificial agents as a novel method for modeling social cognition in multiagent systems. It incorporates an integrated agent control loop to manage, enact and modify frames as models of social interactions. At the same time, it embodies a learning algorithm directed towards social processes with a clear agent- (rather than society-) bias.

6 Conclusion

In this paper, we have proposed interaction frames as a generic data structure for agents who need to manage their interactions with others effectively in dynamical environments among heterogenous, autonomous and possibly purely self-interested agents. Starting from interactionist social theory we have developed a model for interaction processes that focuses on *trajectories* as lines of the interleaving actions of several agents and enriches these with role, context and belief knowledge. By using Goffman's concepts of "frame" and "framing" we laid out an AI-compatible data structure for interaction frames and an agent architecture to manage them.

We believe that these models and architectures provide a new and effective way of dealing with the "social level" of intelligent systems without imposing all to rigid macro-structures that agents have to adhere to.

Still, there is a lot of work ahead: concrete algorithms have to be designed for the adjustment, assessment, matching and update modules of the agent architecture whose effectiveness will have to be evaluated in realistic scenarios. Also, our current theory points at many interesting future extensions which include, but are not limited to,

- developing a theory of *communication and discourse* that is compatible with our current models;
- exploring the usefulness of pre-defined frames (in contrast to the adaptive frames sketched in this paper) as a *software engineering* tool in building multiagent applications;
- exploring the possibilities of framing as a *negotiation* process between agents;
- exploring sociological heuristics to guide the design of concrete computational models of framing;
- analysing the scalability of inter-individual frames to groups and even large societies.

References

[1] M. Barbuceanu and M. S. Fox. Capturing and Modelling Coordination Knowledge for Multi-Agent Systems. *International Journal of Intelligent Cooperative Information Systems*, 5(2–3):275–314, 1996.

[2] C. Castelfranchi. Engineering Social Order. *Proceedings of the Workshop on Engineering Societies in the Agents' World WS at ECAI 2000*, Berlin, 2000.

[3] C. Castelfranchi, F. Dignum, C. M. Jonker, J. Treur. Deliberative Normative Agents: Principles and Architecture. *Proceedings of the Sixth International Workshop on Agent Theories, Architectures, and Languages (ATAL-99)*, 1999.

[4] P. Ciancarini, A. Omicini, and F. Zambonelli. Multiagent systems engineering: the coordination viewpoint. In *Intelligents Agents VI (ATAL99), LNAI.* Springer-Verlag, 2000.

[5] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, (42):213–261, 1990.

[6] P. Cohen and H. Levesque. Teamwork. *Nous*, (35):487–512, 1991.

[7] T. M. Malone and K. Crowston. The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1):87–119, 1994.

[8] K. S. Decker and V. R. Lesser. Generalized partial global planning. *International Journal of Intelligent Cooperative Information Systems*, 1(2):319–346, 1992.

[9] L. Gasser, C. Braganza and N. Herman. MACE: A flexible testbed for distributed AI research. In M. Huhns (ed.), *Distributed Artificial Intelligence*, Morgan Kaufman, Los Altos, CA, 1987.

[10] E. Goffman. *Frame Analysis: An Essay on the Organization of Experience.* Northeastern University Press, 1990.

[11] E. Goffman. *Interaction Ritual: Essays on Face-to-Face Behavior.* Pantheon Books, 1982.

[12] O. Gutknecht and J. Ferber. *Aalaadin: a meta-model for the analysis and design of organisations in multi-agent systems.* Rapport de Recherce, LIRMM, 1997.

[13] A. S. Rao and M. P. Georgeff. An abstract architecture for rational agents. In C. Rich, W. Swartout and B. Nebel (eds.), *Proceedings of Knowledge Represenation and Reasoning (KR&R-92)*, 1992.

[14] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach.* Prentice-Hall, 1995.

[15] M. P. Singh. Toward Interaction-Oriented Programming (poster). *Proceedings of the International Conference on Multiagent Systems (ICMAS)*, Kyoto, Japan, 1996.

[16] M. P. Singh. Multiagent systems as spheres of commitment. *Proceedings of the ICMAS '96 Workshop on "Norms, Obligations, and Conventions"*, Kyoto, Japan, 1996.

[17] M. P. Singh. A Social Semantics for Agent Communication Languages. *Proceedings of the IJCAI Workshop on Agent Communication Languages*, Springer-Verlag, 2000.

[18] M. Wooldridge. *Reasoning about Rational Agents*, MIT Press, 2000.

[19] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter*, MIT Press, 1994.